Perspectival Realism and Frequentist Statistics:

The Case of Jerzy Neyman's Methodology and Philosophy

Adam P. Kubiak

13 September 2022

1. Introduction

Perspectival realism ("PR" hereafter) is a currently developing trend that can be recognised as one of the post-Kuhnian theories of science, within which a remarkable emphasis is put on the fact that cognitive and social dynamics are inseparable elements of the cognitive act and the dynamics of scientific knowledge development (see Collins, Evans 2002). In particular, proponents of PR (premiss1:) "share the general idea that there is no 'view from nowhere' and that scientific knowledge cannot transcend a human perspective", which means the truth condition of a hypothesis, or its justification, depends on an epistemic vantage point, but (premiss2:) "it is in part mind-independent facts that make our theories true or false" (Ruyant 2020).

Scientific outcomes are contingent on the statistical methodology adopted. This non-physical statistical instrument used to design data collection and draw conclusions is prone to a scientist's perspective: there are several possible and acceptable statistical schemes of sampling and inferring, and researchers have to make decisions about the details therein. Therefore, it seems scholarly justified to investigate the interplay between statistics and PR.

It has been argued that PR harmonises with many facts and methodological practices in the formulation and development of scientific theories (see, e.g., Massimi 2018b).

Although PR is sound when applied to cases of substantive content from exact sciences, its relation to statistical methodology appears to be undeveloped. There are perspectival accounts of investigation of aspects of the process of scientific investigation that concentrate on data (see, e.g., Jacoby 2020) observational instruments (see, e.g., Crețu 2020b), and the nature of numerical representations (see, e.g., Wolff 2019). Many authors, like Giere (2010), Rueger (2016), Massimi (2018c) or Potters (2020) argue for perspectivism as regards to scientific (including mathematical) models of experiments and data, but without any specific consideration of the properties of statistical schemes for sampling, inferring and interpretations thereof. These types of analyses are concentrating on substantive contents of models rather than on problems of statistical procedures/tools behind. Giere (1976) discusses frequentist hypothesis testing from a realist position but without a perspectivist aspect fledged. An analysis of the Bayesian statistical approach in the spirit of PR has recently been led (Massimi 2021). The relevant literature falls short in exploring how contemporary PR might interplay with frequentist statistical schemes of sampling and inference. An analysis of relationships between PR and frequentism shall offer a new perspective from which questions about validity, universality, normativity and the philosophical potential of PR can be posed.

The specificity of frequentist statistics (also called classical, orthodox or error-based) is founded on the idea that its inferential scheme lacks probabilistic presuppositions that the state of affairs in question is this or that. The probabilistic elements are introduced in the form of probabilities of detecting the truth were the state of affairs this or that (power function over the domain of possible hypotheses), which pose the basis for knowing pre-observational relative error risks; the power to rightly accept a hypothesis if it's true equals one minus probability of wrongly accepting it if it's false. Another specific element that is particularly important from the perspective of my analysis is that an inferential scheme depends on assumed (under a model of a sampling probability distribution) probabilities of unobserved data.

As far as PR's emphasis on the inseparability of social aspects from the cognitive act is concerned, the important feature of frequentist statistics is that in frequentism social aspects are essentially and explicitly influencing the processes of designing sampling schemes (see, e.g. Kubiak, Kawalec 2022) and setting error risks (see Kubiak et al. 2021). Simultaneously, the method makes it possible to take accepted statistical hypotheses as truly describing a real physical system with objective assessment and control of probabilities of accepting falsehoods (see Giere 1976; Mayo 2018) This brings frequentist methods to be potentially highly interconnected to PR. But frequentism is not homogeneous with respect to methodological and philosophical assumptions (see e.g. Lenhard 2006). Therefore, if PR is to be juxtaposed with frequentism, the proper way is to pick a philosophically and methodologically sufficiently complete version of frequentism offered by a particular author. In the philosophy of statistics, there exists a famous and long-lasting disagreement between Bayesianism and frequentism (see Sprenger 2016). Among frequentists, Jerzy Neyman, with his sharp philosophical views (see, e.g., Neyman 1937, 344; Neyman 1957b), is probably the most historically recognizable opponent of Bayesian statistics. Neyman was a 20th-century statistician who is recognised as one of the co-founders of the frequentist statistical paradigm, which dominated the methodology of natural and social sciences in the 20th century (Lehmann 1985). Re-analyzing Neyman's methodology and philosophy of statistics from a philosophical vantage point that is not pre-determined by the famous controversy between Bayesian and frequentist standpoints shall bring about a new dimension to the debate on his conceptions. As the reader will see, some points in Neyman's writing are balancing between realist and anti-realist (or perspectivist) statements which makes this author's conception even more interesting from the PR's perspective.

My goal is to investigate the extent to which PR is realized in (or can be consistent with) frequentist statistics with an emphasis on the case study of Neyman's conception and to see what consequences for PR can be brought about from such an investigation.

The structure of the article runs as follows. Firstly, in Section 2, I present the PR assumptions (2.1). Based on the problem of the optional stopping rule I offer a motivating example of how PR could fit frequentist statistics in general (2.2). Next, I narrow down the perspective—I reconstruct Neyman's conception of statistical inference with an emphasis on his philosophical views and compare his stance with PR. In Section 3 I discuss aspects in which Neyman's methodological and philosophical views are consistent with realism (3.1) and perspectivism (3.2) and then, in Section 4, I discuss antirealistic (4.1) pragmatistic (4.2) and antipluralistic (4.3) elements of his ideas. Finally, in Section 5 I discuss the issue of genuity of perspectives (5.1) and offer some solutions (5.2-5.4) for problems raised within the three aspects (4.1-4.3). In Section 6 I summarise the results.

2. PR as Applied to Frequentist Statistics

2.1. Assumptions of PR

Perspectival realism is a stance that mediates between the extremes of the objective realist philosophy of science at one end and social constructivist at the other. The perspectivist premiss (1) introduced in Sect. 1 implies that perspectival realism advocates epistemic pluralism (Premiss I): perspectival knowledge about mind-independent states of affairs taken from different points of view can be incompatible, yet still equally valid epistemically because any knowledge of objectively existing facts concerning objects or processes can only be acquired from a perspective (see Massimi 2012).In addition to that, that these perspective-relative claims are true regarding the same objectively existing state of affairs (see the realist Premiss 2 from Sect. 1) implies that (Premiss II) these claims or their justifications retain, cross-perspectively, their performance-adequacy as evaluated from the points of view of the internal standards set by each of the perspectives (see Massimi 2018a, 172); this means the epistemic performance of a scientific claim or a justification (method) must be judged as adequate given standards set by a perspective by practitioners of different scientific perspectives (see Massimi 2018d, 354).

Three general versions of PR can be distinguished. Two refer to the issue of truth-value of statements and in my working classification, they are regarded as semantic-ontological. First, scientific claims can be deemed true relative to a given perspective and "not true simpliciter"(see Creţu 2020a, 1-2). This suggests that although a scientific claim is either true or false as a claim framed within a perspective, the question of its truth-value out of perspective remains meaningless. Second, it can be said that "models are useful to get calculations done but their representational content should not be taken literally as giving us a true story about what the target system is like" because they are about "a modal aspect: it is about exploring and ruling out the space of possibilities in domains that are still very much open-ended for scientific discovery"(Massimi 2018c, 36-38). Model's "being about X is not purported to stand in any mapping relation to worldly-states-of-affairs (X) so as to fulfill the realist quest via a plurality of partially accurate models of X, each of which may give a partial, yet accurate, and veridical image of X." (Massimi 2018c, 38). The third variant shifts the burden of perspectivism from a claim towards its justification(s): "the truthmakers of our beliefs are non-perspectival facts about nature, yet the justification of our beliefs is intrinsically perspectival and rooted in our epistemic perspectives as human agents" (Massimi 2012, 28); this version I call epistemic PR.

The three types of PR are prone to criticism. The first type is criticized for being affected by the problem of relativism: that there are no non-perspectival true claims entails that no non-perspectival facts are illuminated by their meaning, while realism appears to assume that science is (truly) telling what non-perspectival facts are (see Chakravartty 2010). The second type straightforwardly contradicts the very notion of error: it denies the conception of the true value of a parameter being a truth-maker of a rightly asserted true hypothesis presupposed by the idea of the probability of making a right/false assertion. We find the above reasons sufficient not to consider the first two versions of PR in our analysis and to concentrate on the more balanced epistemic PR. The additional premiss (III) to be considered is therefore that there are genuine, non-trivial perspectival justifications (see Massimi 2012). The notion of a perspective is quite vague in the literature and encompasses a type of perspective that could be labeled research traditions as well as narrow perspectives that are sophisticated theoretical frameworks or attitudes of a scientist or group of scientists (see Crețu 2020b). Perspectival aspects of statistical methodology discussed in this paper can be attributed to belonging to both, broad and narrow, categories. On one hand, this methodology encompasses principles or assumptions that form part of the working stance of a scientist, which is classified as a narrow perspective (see Crețu 2020b, 4) but on the other hand, these methodological attitudes are “second-order (methodological-epistemic) principles that can *justify* the scientific knowledge claims advanced” (Massimi 2019, 3) which is classified as a wide perspective (see Crețu 2020b, 29). In this paper, I scrutinise the perspectival nature of a sampling scheme and the inferential pattern both from a general level (e.g. frequentist vs. Bayesian methodological traditions, or approaches) and a detailed level (esp. establishing error risk level or details of observational pattern).

2.2. The Optional Stopping Case-study

An interesting illustration of the PR’s potentiality to encompass the frequentist statistical methodology could be an analysis of the problem of *optional stopping rules*[1] (see, e.g., Savage 1962; Lindley, Phillips 1976) involved in the research example of testing a hypothesis about the sex ratio of the pouch young of koala mothers in poor physical conditions (see McCarthy 2007, 31-33).

[1] The issue of stopping rules as such has been widely discussed by philosophers and methodologists from several perspectives. The case I describe in this subsection is used as an example of how frequentist methodology works, to relate it to PR. The purpose of this paper is not to analyse how this exemplified (or any other) frequentist feature relates to alternative methods, which obey the so-called likelihood principle and by that are insensitive to stopping rules. Nor I intend to discuss wether frequentist violation of that principle is right or wrong. Most generally, the target of this article is analysis of interplay between PR and (Neyman's) frequentism, not analysis of methodological problems or pros and cons of some aspects of frequentist statistics as compared to alternative approaches.

Let us assume that the ecological substantive hypothesis in question states that the proportion of females to males in the population of pouch young is 0.5 (the number of males and females is equal), and the reasonable alternative hypothesis states that the proportion is more than half (females prevail).

The researcher surveyed 12 koala mothers, each with an offspring in its pouch—three of the offspring were males and nine were females. The data could be obtained in at least two ways: the researcher could adopt the experimental rule of terminating sampling once the 12th individual was recorded ($S1$), or the rule to terminate when the 3rd male was recorded ($S2$). Regardless of the sampling strategy, the data seem to be equivalent and the two alternative statistical inferences are as follows.

Sampling in $S1$ is modeled by the binomial distribution that represents the probabilities of collecting n number of females until the number of trials in a sample reaches a fixed value of 12; the sum P_1 of the probability of the observed data (number of females 9) and more extreme data (in this case of having 10, 11, or 12 females in the sample) equals 0,073 thus the observed female ratio in the sample (0.75), given a 0.05 cut-off error rate, is not significantly far from (greater than) the hypothesised population ratio (0.5). The conclusion of the test is not to reject the hypothesis.

Sampling in $S2$ is represented by a different model—the negative binomial distribution that represents the probability of collecting n number of females until the number of males in a sample reaches a fixed value of 3. The p-value P_2 in this case is the sum of the probability of observation and less probable outcomes: having 10 female records, 11, 12, 13, 14, and so on. The p-value equals 0.033 in this case, so with the conventional 0.05 error rate it is significantly low, thus the conclusion is to reject the hypothesis that the population ratio is 0.5.

Therein lies the epistemic oddity—two different sampling strategies, associated with different statistical models of an experiment, lead to different conclusions about the acceptance/rejection of, allegedly the same hypothesis, in the light of, allegedly, equivalent (McCarthy 2007, 37) set of data (evidence) in both cases, consisting of the observation of 9 females and 3 males in a sequence of 12 trials. Below I show how the described frequentist methodological circumstance can exemplify PR.

The two perspectives differ in terms of knowledge claims. This is because they assume different statistical hypotheses: different statements about probability distributions. That is because the sampling spaces and models that serve to formulate the statistical hypothesis tested are different in both cases. The evidence taken into account is also not the same in these two hypothetical cases. The latter is due to both cases assuming a different set of relevant information (evidence) used for inferential purposes. In the case of $S2$, a piece of partial information about the order, i.e. information about the location of the third male record in the sequence of trials is encoded in the (negative binomial) model's random variable. Evidence that was taken into account in $S1$ can be expressed through the proposition: 'exactly three males and nine females were recorded in the sample until (and including) the twelfth trial was recorded in the sample'. In the sampling framework $S2$, the evidence considered can be expressed in the proposition: 'exactly three males had been recorded in the sample until (and including) the twelfth trial and the twelfth trial recorded in the sample was male'. It is easy to see that the second evidence implies the first, but not vice versa, therefore the evidence taken into account is not equivalent for both cases (Kubiak 2014, 138-139).

Despite both possible observational points of view determine two different, perspectival claims and justifications for the conclusions made—what justifies Premiss 1—it is *the same* substantive hypothesis that is being statistically defined and scrutinized in both cases. Both statistical hypotheses are descriptions of a mind-independent state of affairs, an objectively existing population characteristic—the proportion of pouch young females. Common realist feature that represents the claim that the population ratio is 0.5 is one of the models' parameters value (probability of a female in a trial equal to 0.5) represents population ratio. This satisfies Premiss 2.

Epistemic PR concentrates on justification. A statistical procedure/framework can be considered the justification for a conclusion drawn with the use of it (obviously, including the empirical evidence obtained). Therefore, in asking about the performance adequacy of the discussed perspectival statistical set-ups/procedures we are asking about their performance in being epistemically successful perspectival inferential and by that justificatory tools. In both $S1$ and $S2$ the method assumes definite performance adequacy in them being the basis for a true conclusion if the state of the world to which statistical hypotheses defined in $S1$ and $S2$ refer is true. In the considered example with particular data obtained, the conclusions were different in both cases, but this is not inconsistent with PR. The considered method assumes that if the proportion of males in the population of pouch young is 0.5, then the conclusion that will be drawn from $S2$, in the long run, will retain high performance adequacy. That is because if an observation with the use of a sampling strategy from $S2$ were to be repeated iteratively, then the method would anticipate (correct) acceptance of the hypothesis that the population ratio is 0.5 with performance close to the standards set in this method (error risk close to 5%). The same is true for sampling strategy $S1$. Therefore, if the proportion of males in the population of pouch young is 0.5, then both distributions that express the hypothesis tested are true—namely, the value 0.5 of the parameter p is true in both cases of application of different stopping rules and the statistical hypotheses tested in both cases will nominally retain their epistemic performance adequately to standards set for both models; this is a cross-perspectively recognizable methodological fact. This means Premiss II is satisfied. It also appears that the perspectives $S1$ and $S2$, at least in the case considered, where the number of trials and the sex ratio observed are the same, are equally valid epistemically. This means the requirement of epistemic pluralism (Premiss I) is satisfied too.

The above points show that the same objectively existing state of affairs can be scientifically defined via different statistical models that encompass different, incommensurable, observational perspectives and knowledge claims. Although conclusions from testing are to be different for specific evidence possible to be obtained, like the one from the case just considered, the two different ways of framing and testing the substantive hypothesis will have the same performance adequacy if the objective state of affairs represented by the two statistical hypotheses is true. The two ways of scrutinizing the same state of affairs appear to be equally valid epistemically. The upshot is that epistemic PR can have its exemplification in frequentist statistical methodology. In the subsequent two sections of this article, I investigate Neyman's frequentist methodology taken jointly with his philosophical interpretation thereof that can be compared to PR. I start by explicating in Section 3 two elements of Neyman's view that are consistent with PR. Next, in Section 4 I discuss elements inconsistent with PR. Then (Sect 5), in connection with the just analysed example, I draw philosophical consequences for PR.

3. Neyman's Theory—Elements Coherent with PR

Jerzy Neyman was not a professional philosopher; therefore, in communicating his philosophical views he did not use the terminology commonly used in the relevant philosophical debates. Nonetheless, part of his philosophical stance has been explicated and disputed in the philosophical literature (e.g., Hacking, 1965; Mayo Spanos 2006), In this section, I structure those parts of his conception that could be viewed as realism-like and perspectivism-like.

3.1. Neyman's Views and Realism

Some of Neyman's basic methodological and meta-methodological conceptions match realist ideas. Firstly, Neyman did not reject the assumption of the existence of an independent reality (an ontological aspect of realism). He stated that each study involves a "true state of nature" that is unknown (Neyman 1971, 2) this is represented by the *true value* of the hypothesis parameters. The values of the hypothesis parameter(s) that a researcher asks about, were to Neyman "generally unknown constants" (Neyman 1937, 343). The constant value of the statistical model's parameter(s) is as such a mathematical concept, but Neyman writes that “there are real objects that correspond to these abstract concepts in a certain sense” (Neyman 1952, 24). Therefore, the truthfulness of the value of the hypothesis parameter would mean that this value somehow corresponds to, or denotes an unknown, but an independently existing state of affairs in the real world. By that, Neyman seems to be assuming at least ontological realism. What is the nature of the said correspondence?

The general idea of applying statistical schemes to experiments/observations is to “assume that the real value of the sought-after quantity exists […] and—based on laws of large numbers—to seek for calculable measurement results’ functions that can be considered approximations of the ‘true value’ and mean error” (1923a[1], 19, auth. transl.). Therefore, it appears that to Neyman the ideal is to come up with conclusions, in the form of the values of these functions, where “numerical values of mathematical formulas more or less agree with the results of the actual measurements” (Neyman 1952, 24). The values of these functions of actual measurements are expected to be approximately the same as the “real values” that exist independently in the real world, which assumes an epistemic realist approach. What is important from the perspective of further analysis is that Neyman speaks in the plural when he refers to “functions”, which indicates possibly different functions to be used to yield the outcome based on the empirical evidence obtained. Still, all these possibly different outcomes are thought to both agree with the evidence to a certain extent and approximate the objective truth.

The conception of the reliability of the method of statistical inference is anchored in the conceptions of the probability of two types of error: the probability of rejection of the tested hypothesis if it is true and the probability of acceptance of an alternative if it is true (Neyman 1952, 55). A true hypothesis is one in which the stated parameter range covers this unknown, true, real value. The method's reliability is based on performance in yielding true conclusions in the long run. Therefore, a kind of epistemic realism seems to be something that drives the method's reliability in the long term.

Finally, Neyman required the research schemes to be adjusted to the real-world factors that exist objectively and independently of the research scheme. Ignoring these factors might affect the correspondence between a physical (substantial) and a statistical hypothesis. Neyman's illustration of this issue refers to the famous Fisher's toy example of a hypothesis that a lady cannot tell whether the tea or milk was poured into the cup first based on the taste of the tea. An independent factor would be, for instance, an association of the lady's impression of a definite sequence of pouring with the thickness of the cup, which the lady can feel with her lips. If the experiment scheme does not take this into account and it happens that one of the two pouring methods is predominantly used with thinner cups and the other with thicker ones, then the substantial hypothesis of lack of ability may be true, while the corresponding statistical hypothesis—the distribution of probabilities of possible experiment's outcomes under the assumption of lack of ability—will be false (Neyman 1950, 282-291). So the model's (statistical hypothesis') adequacy concerning its describing real-world features is essential when asking about the truthfulness of that model.

It can be concluded that some of the very foundational methodological and meta-methodological conceptions of Neyman appear to match realist ideas. Hypotheses speak of independently existing reality and they are either true or false about this reality. Moreover, the whole research scheme is expected to be adequate with respect to independently existing, real factors. This confirms the presence of some ideas of ontological and epistemic perspectivism in Neyman's thought.

3.2. Neyman's Views and Perspectivism

Although a true statistical hypothesis represents the real value of a quantity existing independently in the world, it does so by rendering the empirical meaning to this real value. A statistical hypothesis is a statement about the probability (density) distribution of a random variable where a random variable is a function of a set of random phenomena obtained in the effect of performing a random experiment. This means the distribution, and so the hypothesis is partially a product of the specificity of the observational (experimental) set-up. Specifically, the notion of probability as used in the statistical hypothesis is that it does not refer to physical objects, or the properties of physical objects, but to the properties of physical events that correspond to an observational setup; in other words, the probability is ascribed not to objects, but events related to an observational setup (Neyman 1952, 10-12). This is visible, for example, in Neyman's comment on Jeffrey's toy example of two boxes. One contains one white and one black ball, while the other has one white and two black ones. Firstly, a box is to be randomly selected, and then a ball at random from that box. Consider Neyman's definition of probability "the probability, $P(B|A)$, of an object A having the property B will be defined as the ratio $P(B|A) = m(B)/m(A)$" (Neyman 1937, 337). When applied to this toy example, it is not the probability of the ball selected having the property of being white:

> "the objects *A* are obviously not balls, but pairs of random selections, the first of a box, and the second of a ball [thus], the probability sought is that of a pair of selections ending with a white ball" (Neyman 1952, 11)

Thus, in the eyes of Neyman, probabilities directly refer to properties of observational designs or procedures. Statistical hypotheses are statements about probability (density) distribution and by that, they are relativised to those designs in the same way. Even if they are to represent substantial hypotheses about the mechanisms or other characteristics of an objectively existing reality, they do so only through the perspectives of experimental constructs that determine what can be experienced and what functions of the data (test statistics) are in use. This view of Neyman's is in line with the consequences of the stopping rule problem discussed in Section 2.

The perspectival nature of frequentist inference is also visible in the socially guided differentiated validity of risks of two types of error (false rejection of a hypothesis called the first type of error and false acceptance of a hypothesis dependent on the power to detect the true alternative): "[…] with rare exceptions, the importance of the two errors is different, and this difference must be taken into consideration when selecting the appropriate test" (Neyman, 1950, 261). From one perspective it might be important to avoid an error of false rejection in the first place and from another "(…) the desirable property of the test of H is as high a power as practicable, perhaps with some neglect of the probability of rejecting H when true" (Neyman, 1971, 4). Consider, for example, a lady who claims to have the ability to distinguish by taste whether milk or tea was first poured in a cup. From the viewpoint of a commission who may grant or refuse a lady's claim of having the ability to distinguish by taste whether milk or tea was first poured in a cup (and perhaps award her for having this distinguishable skill), the more important error to avoid might be granting the claim when in fact it is false. On the other hand, for the lady, the error of falsely asserting that she has no ability might be more important to avoid (Neyman, 1950, 274). Therefore, dependent on the perspective adopted, the error rates might be set in a different way and eventually yield a different decision on whether to reject or accept the hypothesis based on the same statistical test (understood as functions of the data) and the same body of evidence.

The conclusion is that on Neyman's account statistical hypotheses, justifications and conclusions based on the use of statistical tools are always relative to the perspectives of idealised assumptions, experimental constructs and practical perspectives but at the same time refer to the perspective-independent, true states of affairs: real parameter values and real experimental setups and circumstances (see, e.g., Neyman 1934). He thought a fraction of these statements to be true (in the classical sense) to the extent defined by the error risks. Therefore premisses 1 and 2 are satisfied.

Additionally, he appeared to accept the possibility of equally valid perspectives on the foundational assumptions of scientific methodology:

> "[…] in theoretical work, the choice between several equally legitimate theories is a matter of personal taste. In problems of application, the personal taste is again the decisive moment, but it is certainly influenced by considerations of relative convenience and empirical facts" (Neyman 1937, 336 footnote *).

The theories here refer to methodological frameworks. If they can be "equally legitimate", then one can speak of the epistemic pluralism of perspectives, which was told to be an element of PR (Premiss I).

The presented elements of Neyman's conceptions that can be regarded as realism-like and perspectivism-like make his views fairly consistent with perspectival realism thus far. Nevertheless, other important elements of Neyman's approach seem to be inconsistent with PR and also make some of his statements internally inconsistent. These are the theses about the fictional character of scientific concepts, the pragmatistic (non-epistemic) interpretation of a scientific assertion, and the idea of normative anti-pluralism. I discuss these three topics in the follow-up section.

4. Neyman's Theory—Elements Potentially Inconsistent with PR

4.1. Fictional Nature of Scientific Concepts

Due to Neyman statistical hypotheses are stated under idealised assumptions which are false regarding the real world and empirical evidence:

> "The objects in the real world, or rather our sensations connected with them, are always more or less vague and since the time of Kant, it has been realized that no general statement concerning them is possible. The human mind grew tired of this vagueness and constructed a science from which everything that is vague is excluded—this is mathematics. […] there are many mathematical theories that are successfully applied to practical problems. However, this does not mean that these

> theories deal with real objects [...] the theory [of mathematical statistics] itself deals with abstract concepts not existing in the real world" (Neyman, 1952, 23-24).

This might suggest that Neyman believed that there is no truth-correspondence between scientific models and the real world. This seems to explicitly contradict the elements of Neyman's views presented in 3.1.

The problem of the correspondence between the real world, scientific statements and evidence prevail at the level of a single trial and empirical evidence obtained from it. This can be explained by referring to one of Neyman's first papers (1923a[1])[2], where the author introduced a general design for a field experiment conducted for the sake of comparison of different varieties of crops concerning their potential yields.

He considered there the design of the experiment based on the random assignment of seeding to plots in an experimental field. Each seeding ends up with what he called *true yield*. However, the outcome of the measurement of a yield of certain yeast varieties measured (with high accuracy) at a particular plot is not the true yield of that variety at that plot, which is itself an unknown, fixed value (Neyman 1923a[2], 465-67). This divergence is due to the technical error of the measurement. The true yield itself is an idealised conception, namely, the mean value from indefinite repetitions of the measurement with all conditions being equal except for the differences in random technical error that causes inaccuracy of an experimental technique. This kind of error is different than the error in statistical infer-

2 Originally published in Polish. In this article, reference to the Polish original will take the form of "1923a[1]" while the form "1923a[2]" will refer to the fragment translated in 1990 by D.M. Dąbrowska and T.P. Speed.

ence about the hypotheses and "no statistical methods can improve the accuracy of the experiment beyond the limits fixed by the technical random error" (Neyman et al. 1935, 110). Therefore, there is no equivalence between the true yield from a particular trial and an observed yield regarding this trial.

The differences between the two conceptions become striking when one realises that the true yield at a particular trial (plot) is essentially a priori counterfactual state of affairs because of the infinite number of counterfactual unrealised measurements involved in the conception of the true yield (see Rubin 1990). To stress the lack of equivalence between a scientific concept and observable facts, Neyman distinguished two different meanings of terms (such as *yield*) when used in two different aspects of the scientific process: in describing empirical data (Neyman called it "pure empiricism"), and in making inferences to a scientific scheme (Neyman 1923a[1], 18). In the first case, one is speaking of the result(s) of empirical observations (measurements)[3] and in the second—of scientific concepts that put these observations into more general frames. The specificity of using a term in a sense of it being a scientific concept is that "all scientific terms, which are defining properties and relations between investigated objects, are fictions" (Neyman 1923a[1], 18). The *true yield* in Neyman's conception is an example of such a scientific, and therefore fictional expression. It is not clear whether they are fictitious with respect to empirical (potential) observations or real-world properties (mechanisms), or both, but the idea that scientific concepts are fictional somewhat contradicts his realistic views as presented in 3.1 and Premiss 2

---

3 In contemporary apparatus this kind of activity would be labeled *descriptive statistics*, as opposed to inferential statistics.

4.2. The Pragmatistic (Non-epistemic) and Long-run-based Interpretation of Reliability of a Statistical Procedure

The second element of Neyman’s theory possibly inconsistent with PR is the stance that acceptance of a scientific statement does not yield any belief about the truthfulness of a particular scientific statement:

> “The terms ‘accepting’ and ‘rejecting’ are very convenient and are well-established. It is important, however, to keep their exact meaning in mind, and to discard various additional implications which may be suggested by intuition. Thus, to accept [or reject respectively] a hypothesis H means only to decide to take action A rather than action B. This does not mean that we necessarily believe that the hypothesis is true [or false respectively]” (Neyman 1950, 259).

This came to agree with Neyman’s moral-like postulate: “The beliefs of particular scientists are a very personal matter and it is useless to attempt to norm them by any dogmatic formula” (Neyman 1957b, 16).

Epistemic accounts of a hypothesis’ confirmation (or disconfirmation) that are not based on probability or degrees of belief are possible, but Neyman and Pearson stressed that acceptance/rejection of a particular hypothesis based on a statistical test could not—for methodological reasons—be understood as having an epistemic justification. In other words—a single statistical test of a hypothesis does not deliver any measure of the degree of confirmation or disconfirmation of a particular hypothesis:

"[…] as far as a particular hypothesis is concerned, no test based upon the theory of probability can by itself provide any valuable evidence of the truth or falsehood of […] hypothesis. But we may look at the purpose of tests from another viewpoint. Without hoping to know whether each separate hypothesis is true or false, we may search for rules to govern our behaviour concerning them, in following which we insure that, in the long run of experience, we shall not be too often wrong" (Neyman, Pearson 1933, 291).

Nearly twenty-five years later Neyman still stated that statistical tests are decision rules with relative-frequency (long-run) performance, where error probability ascribed to such rule, tells, for example, "How frequently will the contemplated rule prescribe mass application of a given vaccine when, in fact, this vaccine is dangerously toxic?" (Neyman 1957b, 18). The assumption that the method's reliability is not applicable to a single testing situation but to a long sequence of tests implies that it does not serve the role of (dis)confirming particular hypothesis in a single testing situation.

The assumed lack of the epistemic interpretation of a single application of the frequentist procedures is reflected in Neyman's pragmatist interpretation of the goal of the method of scientific investigation with the use of statistical tools. Although "[…] theory was born and constructed with the view of diminishing the relative frequency of errors, particularly of 'important' errors" (Neyman 1977, 108), an acceptance of a hypothesis is an act of will to behave as if the hypothesis was true, based on the assumption that the method, that we are using to do this, is reliable enough not to lead us astray from the truth in a sufficiently large fraction of practically important cases. That is why the final stage of accepting a hypothesis

> "[…] amounts to taking a 'calculated risk', to an act of will to behave in the future (perhaps until new experiments are performed) in a particular manner, conforming with the outcome of the experiment. It is this act of adjusting our behavior to the results of observations, that is the overlooked element of the final stages in scientific research and that is covered by the term 'inductive behavior'" (Neyman 1957b, 12).

Neyman equated his conception of inductive behavior with the decision-theoretic conception of "statistical decision making" introduced by Wald (1950). On grounds of this view one considers expected loss from possible decisions over data probability distribution for a specific hypothesis, where the hypothesis cannot itself be treated as a random variable (see Neyman 1957b; Neyman 1937, 343-344). The choice of the optimal decision rule is based on the analysis of these values for all possible hypotheses. The knowledge about particular decision rule that is required to do that is encapsulated in its "performance characteristic". Consider the function of which value is the probability of making particular decision (e.g. acceptance of the hypothesis tested), and the domain is the set of possible hypotheses (hypothesis' parameter values). Performance characteristic is the system of functions of this type for all possible decisions. One describes the expected losses for all the possible states of affairs and aims first to find the rule that minimizes the worst-case expected loss (es) and secondly to try to minimize expected loss for other (ideally all the rest) possibilities (Neyman 1950, 1-14; 1957b, 18). The described relevant feature of the method of testing is, therefore, pre—observational (it refers to all possible evidential outcomes) and so it does not offer an epistemic interpretation of particular post-observational evidential circumstance and by that the degree of confirmation or epistemic justification of acceptance of a particular hypothesis. It is also essentially pragmatic which blurs its epistemic status even more.

Scientific realism seems to presuppose that scientific conclusions (effects of particular research) are believed to be (rather) true or close to the truth based on statistical justification that gave them a sufficient degree of confirmation. It appears that it would be hard to assimilate epistemic realism with the fact that particular scientific conclusions (outcomes of performing a statistical test) are effects of the need to fulfill pragmatic goals, that they cannot be evaluated in terms of their truthfulness somehow assessed and associated with believing in them being (approximately) true.

4.3. Anti-pluralistic Elements in Neyman's Conception

In Section 3.2, I indicated Neyman's declaration tantamount to the advocacy of pluralism of methodological perspectives. In what follows I show that his methodological solutions and views undermine this supposed pluralism.

Neyman's thought has two anti-pluralistic aspects. Both relate to taking somewhat *God's eye's* view. One is the perspective that could be called the in-theory perspective, and the other is the perspective of justifying the theory from a meta-level point of view. I first discuss the in-theory perspective. It can be further divided into bottom-up and top-down kinds of anti-pluralism. The first is related to the epistemic adequacy of models/setups, while the second refers to the epistemic efficacy of statistical inference.

The bottom-up kind is the one that aims at searching for (selecting) the model of an experiment that is optimally adjusted to physical reality (see Neyman 1950, 282-291). Neyman pinpointed some crucial aspects by referring to the example of the tea-tasting lady (see 3.1). To carry out the experiment, one has to determine an adequate set of admissible hypotheses. For example, should the alternative hypothesis to the one that the lady does not have the ability (she makes random guesses) point in the direction of a perfect guess or perfect misguidance? The lady may be able to discriminate between pouring methods, but simply conflates one method with the other. Another issue is whether one is asking about the lady being able to discriminate between the two methods or identify each. In the second case, the cups should not be judged by comparison between the two in the pair, but independently. However, what if she can identify one of the methods, but is uncertain about the other? Does she know how many cups made with one of the methods she will be given? If so, then the trials should be treated as dependent. Finally, it is essential to arrange a proper technique for a random experiment, in which any factor that may affect the correspondence between a physical and a statistical hypothesis is neutralised by its randomisation, an example of which—referring to the order of pouring—I already provided in 3.1. In conclusion, taking into account manifold factors prompts one to seek the best experimental setup rather than to treat different possible setups as equally valid.

Another example of the bottom-up anti-pluralist approach of Neyman is his theory of utilising sampling design for the sake of optimal estimation (1934; 1938). Speaking of estimation in this context is equally valid as speaking of hypothesis tests, as there is a duality between hypothesis tests and Neyman's technique of estimation by intervals: an estimation technique is tantamount to performing a series of hypothesis tests (see Neyman 1937, 372; Lehmann, Romano 2005,164-168). Neyman develops techniques of how to maximise the accuracy of estimation by taking into account some additional facts about the structure of the population studied in terms of some auxiliary factors. The technique assumes consideration of several, mathematically equally valid, ways of how a sample could be drawn from the population, to choose the one that is, from the perspective of this knowledge, the most accurate sampling design.

All of the above indicates that Neyman advocated achieving the optimal adequacy between the theoretical models of observation and all known aspects of the investigated reality by fulfilling several specific conditions like those above, thus by narrowing down the possible observational perspectives from which a test of a hypothesis could be performed to the one that best corresponds to reality.

The top-down type of Neyman's anti-pluralistic view on the choice of research perspective is perhaps best exemplified by the normative requirement to use a test whereby the probability of correctly rejecting a hypothesis would be maximal for a preassigned error of the first type:

> "if two different critical regions w_1 and w_2 are suggested, both insuring the same probability of error of the first kind, then the choice between these regions de-

pends on their effectiveness in controlling the error of the second kind" (Neyman 1950, 304).

Originally, the rule was presented as applied to choosing among several test statistics (see Neyman, Pearson 1933), but as such, the idea of minimising the error of the second kind can be applied when the choice is to be made between equivalent ways of collecting the data. This is because the critical regions w_1 and w_2 mean two different sets of possible outcomes of observation that lead to rejection of a tested hypothesis if the outcome belongs to a set considered under the assumption of the same risk of the first type of error in both cases. In $S1$ and $S2$ the sets of outcomes that lead to rejection are different.

From this point of view, the two alternative perspectives adopted in the discussed example of testing the hypothesis of the number of males and females of pouch young being even, will not be equally valid epistemically. Imagine a test is run in both cases with the use of the so-called likelihood ratio test statistic; Neyman and Pearson (1933, 298-301) found it to be the statistic that, all things being equal, yields greater power than any other test statistic in the case of point hypothesis testing. Imagine the hypothesis (from 2.2.) that the population ratio is 0.5 is tested against the alternative hypothesis that the ratio in question is 0.75. If the consideration of a test's power function "seems to be the proper rational basis for choosing the test" (Neyman, 1952, 58), then the perspective of sampling design related to $S2$ is preferable. This is due to the likelihood ratio test devised based on sampling distribution $S2$ having higher power (equal 0.46) to detect the true alternative than the power equal 0.39 to detect it in the case of likelihood ratio test devised on the basis of sampling distribution $S1$.

Neyman's meta-methodological views also appear to contrast with the plurality of perspectives by suggesting that, in principle, some methodologies will be more optimal for particular cases than others. Most signifacantly, he admitted that although the Bayesian methodological framework for testing or estimation can be mathematically perfectly valid, it "may be applied in practice only in quite exceptional cases" regarding the conditions of application of the method, namely the lack of evidence to support assumptions about the hypotheses' probabilities: "Even if the parameters to be estimated, $\theta_1, \theta_2, \ldots \theta_l$ could be considered as random variables, the elementary probability law *a priori*, $p(\theta_1, \theta_2, \ldots \theta_l)$, is usually unknown, and hence the [Bayesian formula] cannot be used because of the lack of the necessary data" (1937, 343).

5. Philosophical Consequences

I have argued that PR can have its exemplification in frequentist statistical methodology and be a potentially fruitful explanation of some of its specific features. Nevertheless, when applied to Neyman's frequentism, PR turns out to be only partially consistent with it. Moreover, the sygnalised in 2.1 questionable genuity of perspectives remains unaddressed. In this section, I refer to the issue of the genuine presence of perspectives in the context of the considerations presented to argue for their real, non-trivial presence (5.1). Next, I am checking how far it is possible to solve the problems arising from the comparison of PR with Neymanian conceptions (5.2-5.4).

5.1. Perspectives' Genuity

Epistemic PR (as well as PR in general) is susceptible to reduction to some form of dispositionalism. The argument is that perspectives are reducible to representation/reference to the multi-faceted dispositional nature of the causal properties of the target system. On this account, dispositions are understood as genuinely occurrent real properties that "[d]ispose the systems that have them to behave in particular ways in specific circumstances." (Chakravartty 2010, 409). Dispositions "[…] are non-perspectival facts: they are true whatever perspective one takes. One must take a perspective in order to investigate it, of course; that is, one must view the phenomena from a particular vantage point, or use a particular sort of instrument, or perform a particular kind of experiment, in order to determine how a disposition manifests itself in that particular interaction. But the facts produced by these investigations are perfectly non-perspectival ones" (Chakravartty 2010, 409). So the crux is that evoking different dispositions (most broadly construed) means observing essentially different properties, like in the case of the investigations of the corpuscular and the wave aspects of the light (Chakravartty 2010, 410-411). Similarly, investigations of a material's hardness and its conductivity, or a hornet species' carbon dioxide resistance and its venomousness, would require different experimental interventions and would assume different properties being taken into account. Talking about such dispositions as only perspective-relative real or only perspective-relative knowable does not seem to convey any serious or interesting philosophical content.

Will this argument apply to the features of frequentist statistics? In the Koala example considered it is essentially the same property being observed (the sex of pouch young of randomly selected koala female). This means possibly no causal evocation of different dispositional facts about the population studied. It would be even more bizzare to suggest two different dispositions are evoked in $S1$ and $S2$ when the case of the same observations (including order) in both cases is considered. To be more precise, except for being different in terms of knowledge claims and the evidence taken into account the two aspects are genuinely perspectival also because they are not associated with a difference in observational events encountered in both cases. No causal evocation of different dispositions of the population studied is implied by $S1$ and $S2$: observations in both cases could be assumed to be an identical empirical event (i.e. identical sequence of trial outcomes). Even if one assumes the evidence taken into account in both cases to be different (some formation about the order included in $S2$) it can only be interpreted as caused by the researcher's decision dictated by the choice of the model, not by any ontic difference on the part of the nature of the observed phenomenon. But what about the claim that justification in both cases includes also reference to sets of counterfactual possible absolute (non-relative) frequencies of outcomes, which are different in $S1$ and $S2$, or the claim that dependent on the sampling rule adopted one could encounter different observationational data impossible to be obtainable by the other rule—in the case of S2 the number of investigated koalas could have been larger or smaller. Firstly, different sets of counterfactual unobserved outcomes cannot be said to represent two different 'facts produced' by different dispositional properties, because they do not stand for actually produced (and observed) different facts. Secondly, the difference in possible observations is not qualitative, but quantitative (different number of observed koala mothers), with the feature under scrutiny remaining conceptually the same. There is no way of speaking of essentially different (dispositional) properties involved in $S1$ and $S2$.

Perspectivism present in Neyman (–Pearson's) conception of testing hypotheses is even more striking. While the sampling scheme and raw data may remain the same, the error risks can be set at different levels, or else the data transformations (test statistics) can. These differences mean different conclusion's (decision's) justification content as well as possibly different conclusions driven from the same raw data observed. Can there be anything more perspectival? The Bayesian perspectival priors can match that result. Could there be a stronger—than the one stemming from methodological considerations—case for Bayesian, or frequentist perspectivism in science? Maybe yes, maybe not, but the one presented is already satisfying all the perspectival conditions from the introduction.

And yet is the methodological perspectivism non-trivial enough? It should be noted that the outcomes of random sampling, even under the same model, are by definition also unique (they do not have to include the same individuals).[4] In this sense, any particular observation made by a scientist (or a team) could be interpreted as a unique perspective. Such an understanding of perspective as having its source in the randomness of the process of drawing a sample could be perceived as too trivial to call it perspectival in a philosophically meaningful sense. Yet, the perspectival differences discussed so far are assumed to be anchored in differences in methodological or theoretical assumptions. The 'perspectival' difference in random samples does not result from the application of different observational, inferential or interpretative assumptions. When a scientific methodology is considered,

[4] What is more one researcher (or team) may not have access to samples on another.

perspectives understood as the 'human vantage points' assume some human-driven methodological differences. Coming up with outcomes that are different solely because of the randomness of the process of collecting the data is not human-driven enough in this sense. The differences in the cases of $S1$ and $S2$, as well as differences relating to adopting different testing procedures, esp. error rates, are perspectival in the sense of them having sources in different methodological, pragmatic, and therefore non-trivially human, vantage points. The above considerations confirm Premiss III of genuity and non-triviality of perspectival justifications noticeable in frequentist statistical methodology.

5.2. Overcoming the Unclear Status of Scientific Concepts

I indicate in 4.1. that in discussing the relations between the real world, hypothesis and evidence (data), Neyman pointed at the fictional character of hypotheses (scientific models) and the weak connection between evidence and statistical models. Although the consequence of the latter is that the relation between evidence and independent reality is also weak, this does not mean models (and thus statistical hypotheses) are fictitious when it comes to their relation to independent reality ("real world"). This is because the real world does not reduce to the world of empirical data, although it is "inhabited by data" (Kass, 2011, 2). Neyman only indicated that their fictional nature becomes evident when they are contrasted with "empiricism". In philosophy, empiricism would mean that scientific knowledge reduces to, or is equal to empirical evidence. In this sense, a scientific model clearly states more than what is represented by the outcome of the observation.

If a model of empirical outcome does not conform to a single empirical outcome, this does not mean the model is false or fictitious in terms of it representing the independent reality of more generalised mechanisms and propensities that are the (probabilistic) cause of obtaining an outcome of a certain type. Indeed, Neyman was talking about real, unknown characteristic of a studied population with regard to which a hypothesis as a statement about this characteristic can be true (or false). Empirical data obtained in observation are known, so they must be regarded as something different than those unknown characteristics regarding to which a scientific statement can be true. Therefore, Neyman's statement about the fictionality of all scientific terms should be understood as pertaining to weak evidence—model correspondence, where evidence is something different than the real world. Such a view does not contradict the realistic interpretation of statistical terms and hypotheses concerning real world understood as something different from the world of empirical data. The unknown value of true yield, which can only be estimated, or assumed to be equal to the observed value for practical simplicity, can be representing the propensity of physical (system of) objects to behave in certain, observable ways under repeated observations in certain conditions. Although it is impossible to have an exact empirical realisation of this hypothesized behaviour, the accepted scientific statement can still be regarded as approximately true as regards the unobserved features of the real world.

In 4.1 I indicated Neyman's view that the real world is "vague" while the mathematical models that describe these ontic states are idealisations that deviate from these ontic states and that are "fictitious", which means false. Take, for example, the hypothetical example of a wheel of fortune that was meant by the designer to be fair (which is known to be rather uncommon). The ontic mechanism/character of its propensity must differ from the mathematical model that describes it, as there will never be such a perfect symmetry in it as described by the designer’s model. However, idealisation does not exclude the approximate truthfulness of such models; therefore, realism remains in force. Eventually, it is clear that it is impossible to have completely adequate models, hence Neyman stops short of saying that model must be “found satisfactory” in terms of its empirical adequacy (1952, 27).

5.2. Overcoming The lack of Epistemic Interpretation of a Single Outcome

Neyman seems to be ambiguous as far as realism is concerned, in his insistence that a statistical method is a tool for making pragmatic decisions rather than acquiring true beliefs. This is because he seems to simultaneously assume that the method’s reliability in yielding those pragmatically useful conclusions, in the long run, is based on the method’s reliability in yielding conclusions that are sufficiently often true in a realistic sense, as I argued in 3.1.

Epistemic realism seems to be in force regarding a body of assertions as an effect of the uses of N-P. This is due to error probabilities based on which the procedure may be deemed reliable when it is iteratively used in manifold research contexts with different hypotheses and error rates set at different levels. According to Neyman, the Central Limit Theorem allows us to conclude that the relative frequency of error will be close to the arithmetic mean of the error regardless of context (Neyman 1977, 108-109). This means the average error is an indicator of how a big part of the assertions from a body of outcomes of statistical tests is true, although the question of the truthfulness of any particular one must be abandoned.

It appears then that a special case of realism may apply to Neyman's methodology. This specificity of a realistic interpretation of scientific outcomes would here mean the applicability of realism to them being understood as a collective of outcomes that jointly forms a body of scientific knowledge. A way of arguing for epistemic, and therefore a realistic interpretation of the single-case application of the method in turn could be to show that other measures of a single-case confirmation of a particular hypothesis than just those using the concept of degree of probability or strength of belief can be applied to frequentist methodology. An example of such a frequentist measure that refers to an analysis of power is the conception of "severity" proposed by Mayo and Spanos (see, e.g. 2006): the post-observationally evaluated strength with which an accepted hypothesis has passed the test against definite alternatives; severity can be said to be a measure of the strength of evidential support for a hypothesis in a particular observational circumstance. They even claimed that Neyman also happened to call attention to "post-data use of power" (2006, 334). They showed that on some occasions Neyman wrote vaguely about the confirmation of a hypothesis relative to a high probability of detecting the alternative were it true.[5] This also validates the possibility of a single-case epistemic interpretation of Neyman's method.

5.3. Bogged Down in Neyman's Anti-pluralistic Inclinations?

[5] Nonetheless, he didn't state that the power under consideration was calculated in respect to post-observational p-value, except for one case (1957a, 13) where he spoke of the situation of "large p-value" being obtained.

In the previously cited (in 3.2.) footnote from the 1937 paper, Neyman indicates that when choosing statistical theories (methodological framework) in the context of problems of application, “personal taste” remains “decisive”.In one of the later works, he is even more emphatic by claiming that there should be no “dogmatism” regarding application aspects:

> “What I am opposed to is the *dogmatism* which is occasionally apparent in the application of Bayes’ formula when the probabilities *a priori* are not implied by the problem treated, and the author attempts to impose on the consumer of statistical methods the particular *a priori* probabilities invented by himself for this particular purpose” (Neyman 1957b, 19).

Methodological choices can determine the outcome of the application of statistical procedures.[6] However, “inventing” the prior also determines the outcome, and can also be understood as a methodological choice based on “personal taste”. Why influential personal “choice”, “taste”, “invention”, you name it, is acceptable when Neyman speaks of the statistical methodology adopted by him and is not acceptable when Neyman speaks of the Bayesian statistical methodology remains unexplained and seems inconsistent.

As I illustrated in 4.3, there is a tension between the idea of the decisiveness of “personal taste” in choosing the mathematical construct, which entails the acceptance of the pluralism of perspectives, and the tendency to eradicate the equivalence of perspectives both at the methodological and meta-methodological level. I have shown it in 4.3. that ac-

6 I exposed this in Section 2.

cording to Neyman the epistemically best perspective should be determined by the objective context of the research; this could be for example prior knowledge of the population studied, or of conditions of the experiment. The antipluralistic meta-methodological tendency is especially visible in the fragment (paradoxically, from the same paper where the mentioned footnote is) cited at the very end of 4.3. There he suggests that the choice of a statistical framework should be tied to physical reality. Despite all that Neyman clearly stated that there is an aspect of his methodology where pluralism of perspectives (methodological choices) is allowed.

The decisive, pluralistic element in Neyman's conception is explicitly present in the pragmatic choice of standards for the two types of error. Despite Neyman believed that "this subjective element lies outside of the theory of statistics" (Neyman 1950, 263) it determines an important part of the statistical procedure and outcomes so Neyman's method can be regarded as advocating perspectival pluralism in respect of the pragmatic-value driven differentiation of error risks. Does this perspectival aspect make different perspectival statistical justifications potentially equally valid epistemically? Consider, for example, two hypothetical settings of error risks for a research question. The first with the risk of falsely rejecting the hypothesis tested equal 5% and of falsely accepting it equal 20%. The second is with false rejection risk at the level of 20% and false acceptance risk—at 5%. For some observations the two test settings would give different results. If the hypothesis tested is true, the epistemic performance of the first procedure will be greater than that of the second. But, if the alternative is true the epistemic performance of the second will be greater by the same amount. As it was argued, the default Neymanian approach is to understand the reliability of the procedure as a pre-observational feature with both types of errors being relevant for the assessment of it. In addition to that, no prior information about the possibility of any hypothesis being true is applicable. Because of that if only the epistemic aspect of the set error risks—without consideration of their pragmatic rationale—is taken into account, the two testing settings described must be regarded equally valid epistemically. This means Neyman's conception implicitly assumes partial epistemic pluralism (Premiss I is partially satisfied) despite of Neyman's legitimate impulse to weed out the subjective, pluralistic element from his methodology.

However, this does not solve the problem of Neyman's inconsistency in him criticising Bayesians for their pluralism of setting the priors. Both setting error risks in frequentism and setting the prior probability of a hypothesis in Bayesianism are factors that influence the outcomes of the application of the respective procedure. The choice of which type of error is more important (and to what extent) than the other can be—according to Neyman—burdened with subjective, personal taste, but the choice of the prior cannot. Therein lies the inconsistency: he banishes Bayesianism for the presence of subjective element but allows for the presence of subjective element in his method. It appears that the problem boils down to the question of what can be assumed to be "outside of the theory of statistics", and what may be thought to belong to it. Neyman seemed to push pluralism aside to this outer world and decide that setting the nominal value of the error rate is of this outside-world type while setting the value of the prior probability would be of the inside-world type. Yet, he did so in a somewhat arbitrary way.

A tactic to excuse Neyman could be to say that Neyman's dissatisfaction with Bayesian priors, while accepting a role for preferences in error risk management, could be due to the fact that Bayesian priors are not supposed to be value-laden (since they are degrees of credence, and since epistemic preferences are not allowed). Error risk considerations can in turn be value-laden, and subject to preferences that, at least to Neyman, remained on the pragmatic, not epistemic level.[7]

6. Conclusions

In this article, I, firstly, argued that PR appears to be a possible framework for frequentism. To do that I used the example of stopping rules. Secondly, I made an extensive analysis of how Neyman's frequentism relates to the set of PR's basic assumptions. It turned out that there are some aspects of Neyman's thought that seem to confirm PR, and others that disconfirm it. Eventually, it was possible to show that epistemic PR is consistent with Neyman's frequentism to a satisfactory degree. I have shown that on the grounds of Neyman's frequentist methodology one is dealing with genuine—irreducible to dispositional explanations—and non-trivial perspectives. The problematic aspect of the alignment between PR and frequentism is that on Neyman's account epistemic pluralism of perspectives seems to be discredited. This is especially noticeable in the fact of the unequal epistemic status of the two stopping rules analysed, and in Neyman's methodological and meta-methodological statements and solutions aiming at finding, among several possible

---

[7] I thank to the anonymous Reviewer for drawing attention to this interesting route of inquiry.

frameworks, the ones that are most optimal epistemically. Nevertheless, the fact of the possibility of setting the rates of two types of errors in different ways will, in some cases, lead to epistemic pluralism of methodological perspectives. This is one aspect of Neyman's frequentist conception in which perspectival pluralism may be the case but perspectival pluralism, in general, is not a normative principle and should be avoided whenever it is possible. This means Neyman does not postulate eliminating all plurality, but narrowing down the number of possible perspectives as much as we can, with the possibility of remainings of some residual pluralism. For this reason, Neyman's methodology both fits and partly undermines PR by making it a descriptive rather than normative position, and by making it a case (or aspect)-dependent instead of universal, absolute, or binding stance.

References

Chakravartty, Anjan. 2010. "Perspectivism, inconsistent models, and contrastive explanation" *Studies in History and Philosophy of Science Part A*, Volume 41 (4), 405-412.

Collins, Harry M., Evans, Robert. 2002. "The third wave of science studies: Studies of expertise and experience." *Social Studies of Science* 32: 235–296.

Crețu, Ana-Maria. 2020a. "Perspectival Realism." In *Encyclopedia of Educational Philosophy and Theory*, ed. M. A. Peters. Singapore: Springer.

——— 2020b. "Perspectival Instruments" [preprint] accessed 06.04.2021 http://philsci-archive.pitt.edu/id/eprint/18341

Fisher, Ronald A. 1956. *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd.

Giere, Ronald N. 1976. Empirical Probability, Objective Statistical Methods, and Scientific Inquiry. In Harper, W.L., Hooker, C.A. (eds) *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science. The University of Western Ontario Series in Philosophy of Science*, vol 6b. Springer, Dordrecht.

——— 2010. "An agent-based conception of models and scientific representation." *Synthese* 172: 269-281.

Hacking, Ian. 1965. *Logic of Statistical Inference*. London: Cambridge University Press.

Jacoby, Franklin. 2020. "Data identity and perspectivism." *Synthese* (2020). https://doi.org/10.1007/s11229-020-02824-8

Kass, Robert E. 2011. "Statistical Inference: The Big Picture." *Statist. Sci.* 26 (1): 1-9.

Kubiak, Adam P. 2014. "A Frequentist Solution to Lindley & Phillips' Stopping Rule Problem in Ecological Realm." *Zagadnienia Naukoznawstwa* 50 (200): 135-145.

——— and Kawalec, Paweł. 2022. Prior Information in Frequentist Research Designs: The Case of Neyman's Sampling Theory. J Gen Philos Sci (2022). https://doi.org/10.1007/s10838-022-09600-x

——— Kawalec, Paweł and Kiersztyn, Adam. 2022. „Neyman-Pearson Hypothesis Testing, Epistemic Reliability and Pragmatic Value-Laden Asymmetric Error Risks." Axiomathes 32, 585–604.

Lehmann, Erich L. 1985. "The Neyman-Pearson Theory After Fifty Years." In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Volume*

1, ed. L.M. Le Cam, 1047–1060. Michigan: Wadsworth Advanced Books & Software.

Lehmann, Erich, L., Romano, Joseph, P. 2005. *Testing Statistical Hypotheses*.Third Edition. New York: Springer-Verlag.

Lenhard, Johannes. 2006. "Models and Statistical Inference: The Controversy between Fisher and Neyman–Pearson." *The British Journal for the Philosophy of Science* 57: 69–91.

Lindley, Dennis V., Phillips, Lawrence D. 1976. "Inference for a Bernoulli Process (a Bayesian View)." *The American Statistician* 30 (3): 112–119.

Massimi, Michela. 2021. "Cosmic Bayes. Datasets and priors in the hunt for dark energy." *Euro Jnl Phil Sci*, 11, 29 (2021). https://doi.org/10.1007/s13194-020-00338-1

——— 2019. "Realism, perspectivism, and disagreement in science." *Synthese* (2019). https://doi.org/10.1007/s11229-019-02500-6

——— 2018a. "Perspectivism." In *The Routledge Handbook of Scientific Realism*, ed. J. Saatsi, 164-175. Oxford: Routledge.

——— 2018b. "Three problems about multiscale modeling in cosmology" *Studies in History and Philosophy of Modern Physics,* 64**:**26–38.

——— 2018c. "Perspectival Modeling." *Philosophy of Science* 85(3): 335–359.

——— 2018d. "Four kinds of perspectival truth." *Philosophy and Phenomenological Research*, 96(2), 342–359.

——— 2012. "Scientific perspectivism and its foes." *Philosophica* 84(1): 25–52.

Mayo, Deborah. 2018. *Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars*. Cambridge: Cambridge University Press.

——— and Spanos, Aris. 2006. "Severe Testing as a Basic Concept in a Neyman-Pearson Philosophy of Induction." *The British Journal for Philosophy of Science*, 57 (2): 323–357.

McCarthy, Michael A. 2007. *Bayesian Methods for Ecology*. Cambridge: Cambridge University Press.

Neyman, Jerzy. Spława-Neyman, Jerzy. 1923a[1]. "Próba uzasadnienia zastosowań rachunku prawdopodobieństwa do doświadczeń polowych." RocznikiNaukRolniczychi, Leśnych 5: 1–51.

——— Spława-Neyman, Jerzy. 1923a[2]. "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." Part of (Neyman 1923a[1])transl. by T. M. Dabrowska, T. P. Speed. Statistical Science 5: 465–472.

——— 1934. "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection." *Journal of the Royal Statistical Society* 97 (4): 558–625.

——— 1937. "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability." *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 236 (767): 333–380.

——— 1938. "Contribution to the Theory of Sampling Human Populations." *Journal of the American Statistical Association* 33:101–116.

——— 1950. *First Course in Probability and Statistics*. New York: Henry Holt and Co.

——— 1952. *Lectures and conferences on mathematical statistics and probability*. Washington: U.S. Department of Agriculture.

——— 1957a. "The Use of the Concept of Power in Agricultural Experimentation." *Journal of the Indian Society of Agricultural Statistics* 9: 9–17.

——— 1957b. "'Inductive Behavior' as a Basic Concept of Philosophy of Science." *Revue De L'Institut International De Statistique* 25 (1/3): 7–22.

——— 1971. "Foundations of Behavioral Statistics." In *Foundations of Statistical Inference,* ed. V. P. Godambe and D. A. Sprott. Toronto: Holt, Rinehart and Winston.

——— 1977. "Frequentist probability and frequentist statistics." *Synthese*36 (1): 97–131.

——— and Pearson, Egon, S., 1933. "On the Problem of the Most Efficient Tests of Statistical Hypotheses." Philosophical Transactions of the Royal Society of London. Series A 231: 289-337.

Potters, Jan. 2020. "Perspectivism and the epistemology of experimentation." Euro Jnl Phil Sci 10, 24. https://doi.org/10.1007/s13194-020-00289-7.

Sprenger, Jan. 2016. "Bayesianism vs. Frequentism in Statistical Inference. In The *Oxford Handbook of Probability and Philosophy*, ed. A. Hájek and C. Hitchcock, 382-405. Oxford: Oxford University Press

Rubin, Mark. "Repeated sampling from the same population?" A critique of Neyman and Pearson's response to Fisher. *Euro Jnl Phil Sci* 10, 42 (2020). https://doi.org/10.1007/s13194-020-00309-6

Rueger, Alexander. "Perspectival Realism and Incompatible Models." *Axiomathes* 26: 401–410.

Ruyant, Quentin. 2020. Perspectival realism and norms of scientific representation. Euro-*Jnl Phil Sci* 10, 20 (2020). https://doi.org/10.1007/s13194-020-00285-x

Savage, Leonard. ed. 1962. *The foundations of statistical inference: A discussion.* London: Methuen.

Wald, Abraham. 1950. *Statistical Decision Functions*. New York: John Wiley and Sons.

Wolff, J.E. 2019. Representationalism in Measurement Theory. Structuralism or Perspectivalism? In *Understanding Perspectivism: Scientific Challenges and Methodological Prospects*, ed. M. Massimi& C. D. McCoy, 109-126. London: Routledge.